

JSC370 2022: Midterm Project

Due Thursday, March 3rd by 11:59pm Eastern Time

Learning Objective To apply the skills learned in JSC 370 (up to February 17th) by analyzing and interpreting a dataset of your choice.

Narrative Through this project you will launch a portfolio of data science projects that will become seminal for your job hunt. This midterm is a stepping stone for the final project. The first step in any data analysis is to have a dataset for which you have formulated an interesting question. If you do not have a dataset to work with, you may choose one from our list of suggestions. With your dataset, formulate a clear and concise question to answer and conduct exploratory data analysis, and data visualization to explore/answer this question.

Deliverable: A knitted R markdown *written report* (.html or .pdf) with embedded tables and figures that is submitted to a project-specific github repository that you create. The report should have the following sections:

- Introduction (provide background on your dataset and formulated question),
- Methods (include how and where the data were acquired, how you cleaned and wrangled the data, what tools you used for data exploration),
- Preliminary Results (provide summary statistics in tabular form and publication-quality figures, take a look at the kable function from knitr to write nice tables in Rmarkdown),
- Conclusion/summary about what you found in terms of the formulated question.

In your report, please *do not* include unformatted output or dataset summaries (e.g. output from `head()`, `str()`, etc.). You should summarize these aspects of your data within the text.